Consulting Psychology Journal: Practice and Research
2010, Vol. 62, No. 4, 216–234

1065-9293/10/$12.00 DOI: 10.1037/a0022265

HOW TO (AND HOW NOT TO) ASSESS THE INTEGRITY OF MANAGERS

Robert B. Kaiser
Kaplan DeVries Inc.

Robert Hogan
Hogan Assessment Systems

This article concerns how to evaluate the integrity of managers, with an emphasis on identifying those with low integrity. After defining our terms, we review leadership research showing that subordinate perceptions of their manager's integrity determine how much they trust their manager which, in turn, influences their attitudes and performance. Next we evaluate a common method for assessing the integrity of managers. This method defines integrity as a leadership competency and measures it using coworker ratings of observed ethical behavior. We found that these behavioral ratings suggested only a negligible proportion of managers may have integrity issues and did not distinguish high- from low-performing managers. We then propose an alternative method based on subordinate expectations about the likelihood that their boss would behave unethically. This method suggested that a much larger proportion of managers may have integrity issues and did distinguish high- from low-performing managers. Based on these findings, we offer recommendations for assessing integrity in practice and urge the leadership field to seriously consider the prevalence and impact of managerial misconduct.

Keywords: integrity, ethics, leadership

Character is personality evaluated; personality is character devaluated.
— Gordon Allport

In her timely book, *Bad Leadership*, Barbara Kellerman (2004) pointed out the curious tendency of the field to glorify leaders while overlooking the harm they sometimes do. But as events of the last decade have made clear—beginning with the criminal acts that toppled Enron and Tyco in 2001 through the unscrupulous behavior leading to the global economic meltdown of 2008–2009—some people in leadership roles behave badly. Beyond the sensational examples, daily organizational life includes regular episodes of staff abuse, theft, rule bending, and skullduggery perpetrated by people in positions of authority. These occurrences can be traced to issues of integrity (Padilla, Hogan, & Kaiser, 2007; Thompson, Grahek, Phillips, & Fay, 2008).

This article concerns how to evaluate the integrity of managers, with an emphasis on identifying those with low integrity. We begin by defining our terms; we next review research on the role of integrity in leadership. Then we review some suggestions from personality theory regarding how to assess integrity. Next, we evaluate a common method for measuring integrity that defines it as a leadership competency and uses coworker ratings to measure it. We report an empirical study

Robert B. Kaiser, Kaplan DeVries Inc., Greensboro, North Carolina; Robert Hogan, Hogan Assessment Systems, Tulsa, Oklahoma.

Correspondence concerning this article should be addressed to Robert B. Kaiser, Kaplan DeVries Inc., 1903-G Ashwood Court, Greensboro, NC 27455. E-mail: rkaiser@kaplandevries.com

showing that this method identifies few managers with potential integrity issues and does not distinguish high- from low-performing managers. We then propose an alternative method based on subordinate expectations regarding the likelihood that their managers would misbehave. Next we report a study of the ability of this method to identify managers with potential integrity issues and predict indices of effective leadership. Finally, we discuss the practical implications of our analysis and recommend that the leadership field pay greater attention to managerial misconduct.

Definitions

Integrity and Related Concepts

One can distinguish between integrity, ethics, morals, and character, but the way these terms are normally used suggests they are similar. For example, the *Oxford Dictionary* contains the following definitions: (1) integrity (n.d.), "the quality of being honest and having strong moral principles;" (2) ethics (n.d.), "moral principles that govern a person's or group's behavior;" (3) morals (n.d.), "standards of behavior or beliefs concerning what is and is not acceptable to do;" and (4) character (n.d.), "a person's good reputation." Because these terms are used to define each other, they are roughly synonymous (cf. Ciulla, 2004). We use the terms integrity, ethics, and character interchangeably in this article.

The study of ethics is the traditional province of philosophers. Ethics refers to the values and behaviors that society defines as desirable and that provide the rules for judging actions as "good or bad" (Pojman, 1995). Northouse (2006) noted two types of ethical theories relevant to leadership, those that focus on conduct and those that focus on character. From the conduct perspective, the actions of a leader can be judged by their consequences—for instance, Bentham's utilitarianism favors actions that produce the "greatest good for the greatest many"—or by their compliance with moral rules—for example, Kant's view of duty as the obligation to do the right thing. From the character perspective, a leader's conduct is determined by the kind of person he or she is. For example, Aristotle defined maturity in terms of developing a good character which comes from practice (e.g., by telling the truth, one develops the virtue of honesty). Ethical behavior, in other words, is a developed disposition or character trait.

Most moral philosophies assume that ethics and integrity concern one's relationships with other people. Violations of ethics typically involve harming others (Ciulla, 2004; Northouse, 2006). Lying, cheating, and stealing represent attempts to advance one's self-interest while ignoring the rights of other people. Conversely, leaders who display integrity seem concerned about the welfare of others (Brown & Trevino, 2006).

The term *integrity* can be used in two distinct ways, and each requires clarification. The first refers to honesty; although this involves playing by the rules, it is different from following the rules—it is more like "fair play." People sometimes use rules in order to harm others, whereas fair play may require setting the rules aside in cases where people will be victimized by them. Further, analyses of recent corporate ethical failures note that although certain practices were within the letter of the law, they were nonetheless unethical (Jennings, 2006).

According to the *Oxford Dictionary*, the second use of the term integrity (n.d.) is "the state of being whole and undivided"—for instance, consistency between words and actions. We think integrity is best understood as a moral attribution that we place on another person's conduct rather than a statement about the consistency of that person's words and actions. Adolph Hitler illustrates this point: as the German Chancellor, there was "wholeness" and "unity" between his anti-Semitic rhetoric and his political actions. But most observers judge his behavior immoral. Like beauty, integrity is in the eyes of the beholder.

Leadership

Our view of leadership differs from the view of leadership as a position, which assumes that if someone is in charge of something, then by definition that person is a leader. In contrast, we view leadership from the perspective of human origins and see it as a resource for group survival (Van

Vugt, Hogan, & Kaiser, 2008). The solution to most of the survival problems faced by early humans required collective action—to hunt large game, to ward off predators, to repel invading tribes. Consequently, we think of leadership as an adaptive solution to the problem of coordinating collective effort. We believe that leadership emerged as a mechanism for influencing individuals to transcend their short-term selfish interests and work together for the long-term welfare of the group. In this view, leadership involves building a team and guiding it to outperform its competition (Hogan, Curphy, & Hogan, 1994; Hogan & Kaiser, 2005; Van Vugt et al., 2008).

Integrity and Leadership

Peter Drucker argued that abiding by business ethics and displaying personal integrity are prerequisites for leadership; although ethics and integrity may not ensure effective leadership, Drucker claimed that their absence precluded it (Cohen, 2009). Research on implicit leadership theories, organizational trust, transformational leadership, leader-member exchange, and the emerging evolutionary perspective all support this claim. The findings indicate that leaders need to be viewed as having high integrity in order to win the trust of followers, and when leaders are seen as lacking integrity, it harms the trust and relationships needed to build and maintain a team.

Research on implicit leadership theories suggests that people have innate cognitive categories that they use to evaluate another person's leadership potential (Lord & Maher, 1993). This work shows that the single most important evaluative criterion for deciding whether someone is worth following or not is "honesty," although "fair" and "believable" are also high on the list (Lord, Foti, & De Vader, 1984). Consider, for example, two of the most respected U.S. presidents: Abraham Lincoln is remembered as "honest Abe," and George Washington is said to have "never told a lie." Further, the centrality of honesty in perceptions of leadership applies around the world. The GLOBE study of leadership in 62 cultures found that "trustworthy," "just," and "honest" were universally desired attributes (Den Hartog, House, Hanges, Ruiz-Quintanilla, & Dorfman, 1999). In fact, across all cultures, "trustworthy" was the attribute that followers rated as most important for effective leadership.

Research on organizational trust emphasizes the importance of managerial trustworthiness (Shockley-Zalabak, Ellis, & Winograd, 2000). A meta-analysis of this literature shows that the degree to which employees trust their direct supervisor is correlated with job satisfaction, job performance, and exercising discretionary effort (Dirks & Ferrin, 2002). There were also effects for trust in management in general, but they were small compared to the effects of trust in direct supervisors. The authors speculated that trust in one's boss may be the strongest determinant of employee outcomes in the workplace.

Research on transformational leadership also highlights the role of integrity. Transformational leadership emphasizes a collective vision, considers the needs of individual followers, and helps them align their self-interests with collective interests. In his original formulation of the concept, Burns (1978) emphasized moral development. Bass and Steidlmeier (1999) argued that transformational leadership depends on the character of leaders, the values reflected in their visions, and the morality of their methods. Indeed, followers' perceptions of transformational leadership are closely related to their perceptions of a leader's integrity (Parry & Proctor-Thomson, 2002). Conversely, integrity seems not to be central to charismatic leadership. In fact, researchers distinguish narcissistic charismatics who seek self-advancement from prosocial charismatics who put the interests of the organization ahead of themselves (Howell & Avolio, 1992; O'Connor, Mumford, Clifton, Gessner, & Connelly, 1995).

Leader-member exchange (LMX) theory proposes that, in effective groups, leaders and followers develop relationships based on mutual trust and respect (Dansereau, Graen, & Hagan, 1975). Meta-analysis confirms that the stronger the relationship, the more positive are follower outcomes such as motivation, commitment, and performance (Gerstner & Day, 1997). LMX theory is derived from social exchange theory (Blau, 1964), which maintains that people remain in relationships only if they feel what they receive from a relationship is proportional to what they put into it. Subordinates' sense of being treated fairly is the core of the concept of engagement. When subordinates think they are not treated fairly, they will get even—by withholding effort, by theft and sabotage, or by leaving the relationship and/or the organization (Greenberg, 1990).

Finally, an evolutionary perspective on leadership brings these various lines of research together by emphasizing two points (cf. Van Vugt, Hogan, & Kaiser, 2008). First, leadership is best understood by considering the costs and benefits for both leaders and followers. The benefits associated with status make it obvious why some people want to lead. However, it is not obvious why anyone would want to follow. The reason seems to be that although not everyone can be the leader, followers in well led groups are better off than those in poorly led groups (Van Vugt et al., 2008). Therefore, it pays to follow good leaders and is costly to follow bad leaders.

The second point concerns the fundamental ambivalence of leader–follower relationships. On the one hand, followers depend upon leaders for collective success. On the other hand, leaders are often tempted to exploit their position for personal gain (Maner & Mead, 2010; Van Vugt et al., 2008). Dominance and despotism are part of our primate heritage, and as Freud (1921) suggested, people who aspire to positions of power are often selfishly motivated. Not surprisingly, there is evidence that evolved mechanisms help followers avoid being taken advantage of by selfish leaders. For example, in all primate groups, members pay disproportionate attention to higher-status individuals (Chance, 1967). Among humans, gossip and public criticism are often used to control bad leaders (Boehm, 1999). Humans also have finely-tuned mechanisms for detecting individuals who take more than they give in relationships and cooperative efforts (Cosmides & Tooby, 1989, 1992). Strong instincts for reciprocity, which Piaget (1965) observed among young children, also encourage followers to avoid or sanction unfair leaders. These findings help explain why followers in every culture are concerned about the trustworthiness of leaders and why perceived integrity plays such a pivotal role in leader–follower relations. At a deep, unconscious level, followers are wary of leaders who may abuse their power.

Personality Theory and the Measurement of Integrity

Gordon Allport (1937) and the moral philosophers agreed that issues of integrity are linked to personality. Allport explicitly stated that character reflects an evaluation of one's personality. Similarly, whether philosophers analyze principled behavior, the impact of behavior on other people, the motives for behavior, or the virtues displayed in behavior, they are also concerned with evaluating characteristic patterns of social behavior.

There is a connection between personality psychology and integrity testing. The integrity tests commonly used for preemployment screening do not measure integrity directly; rather, they measure a combination of the three dimensions of the Five-Factor Model of personality that reflect socialization: Conscientiousness, Agreeableness, and Emotional Stability (J. Hogan & Ones, 1997). Although these tests predict counterproductive work behavior (Ones, Viswesvaran, & Schmidt, 1993), they do not measure counterproductive behavior. Moreover, there is a question about how well integrity tests work with leaders (Howard & Thomas, 2010). As the head of assessment for a major executive search firm put it, "I have not found (integrity questionnaires) to be frequently used for senior executives. Many of these existing integrity questionnaires are more "junior" in their focus and attend to items such as stealing staplers and other office products" (Stamoulis, 2009, p. 92).

Our approach to using personality psychology to assess integrity does not rely on the adoption of established methods for measuring personality. Rather, we use personality theory to identify optimal conditions for measuring the integrity of managers. Our approach draws on the concepts of reputation, the dark side of personality, and the influence of "weak" situations on the expression of dark-side tendencies.

Identity Versus Reputation

MacKinnon (1944) observed that personality should be defined in two ways. On the one hand, personality refers to factors inside people that explain their behavior. Different theorists prefer different internal factors—egos, temperaments, schemas, and so on. We call this personality from the perspective of the actor. On the other hand, personality refers to the distinctive impressions that people make on others, impressions which are captured in the adjectives we use to describe others (friendly, honest, principled, etc.). We call this personality from the perspective of the observer. The first definition of personality concerns how people think about themselves—their "identity," and the second definition concerns how others think about them—their "reputation" (Hogan, 2007).

The best predictor of future behavior is past behavior; reputation is a summary of past behavior and is, therefore, the best data source we have about future behavior. This assertion is supported by a wealth of empirical data (see Hogan, 2007). In contrast, identity is difficult to measure because self-perception is notoriously plagued by bias and error. Self-ratings of behavior have low correlations with both observer ratings and objective measures (Beehr, Ivanitskaya, Hansen, Erofeev, & Gudanowski, 2001; Heidemeier & Moser, 2009). There is also a logical problem with self-assessments of integrity. People who lack integrity specialize in manipulation and deceit, which makes their self-assessment a dubious source of information.

Bright Side Versus Dark Side

It is also useful to distinguish two aspects of reputation, "the bright side" and "the dark side" (Hogan & Hogan, 2001). The bright side reflects people's reputations when they are on their best behavior—when they self-monitor and self-regulate to make a positive impression (e.g., during an employment interview). The dark side refers to people's behavior when they are less concerned about how they are perceived—when they let down their guard or when they are too stressed or too tired to self-monitor. The bright side concerns people when they are at their best; the dark side concerns people when they are less vigilant.

Dark-side characteristics can be seen as dysfunctional extensions of bright-side characteristics. For instance, extreme self-confidence resembles narcissism and extreme flexibility resembles delinquency (Hogan & Kaiser, 2005). Further, dark-side characteristics often coexist with well-developed social skills that mask their dysfunctional impact in the short-term (Hogan & Hogan, 2001). However, over time and with repeated exposure, observers come to recognize these dysfunctional tendencies that disrupt relationships and corrupt judgment (J. Hogan, Hogan, & Kaiser, 2010).

Strong Versus Weak Situations

The personality critic Walter Mischel (1977) suggested that personality is most important in "weak situations." According to Mischel, "strong situations" provide unambiguous cues about appropriate behavior, which reduce the variability in peoples' actions. In strong situations, clear behavioral cues encourage people to conform to social expectations. Conversely, weak situations provide ambiguous cues for action; this allows greater opportunity for personality to influence behavior. Moreover, because weak situations lack cues for appropriate behavior, socially undesirable behavior is more likely to be expressed in them. Therefore, the dark side of personality is more likely to appear in weak situations (Kaiser & Hogan, 2007).

Hierarchical relationships can be understood in these terms—relations with superiors represent strong situations and relations with subordinates represent weak situations. It follows that managers are more likely to be on their best behavior when dealing with their superiors, but less likely to be as well-behaved in the weaker situations involving their subordinates. Thus, subordinates may be in the best position to observe a manager's dark side. And given the vulnerability of their subordinate position, they also should be motivated to detect signs of exploitation and untrustworthy behavior. These factors make subordinates a prime source of information about the integrity of their boss.

Implications

These considerations have three implications for the assessment of integrity. First, managers who lack integrity probably won't describe themselves that way. Second, observer evaluations should be better at identifying managers with questionable integrity. And third, subordinates are likely to be the best source of information about a manager's integrity. This conclusion converges with recommendations made by leading researchers of leadership and ethics (cf. Brown & Trevino, 2006; Craig & Gustafson, 1998).

Ratings are the most common way to assess performance in organizations (Murphy & Cleveland, 1995). The next section examines how subordinate ratings are used to evaluate the integrity of managers in the context of competency models and 360-degree feedback surveys. These ratings focus on behaviors that indicate the presence of integrity. We highlight the problems with this method using data from executives in a Fortune 500 firm. Then we consider a method that focuses on subordinates' expectations regarding the behavior of their managers, an approach that seems to overcome the limitations of competency ratings.

Study 1: Competency Ratings of Integrity

Most organizations use competency models to identify the performance capabilities that distinguish high- from low-performing managers (Boyatzis, 1982). Competency models frequently include dimensions like integrity, although they may be labeled something else—for example, credibility, ethics and values, integrity and trust, trustworthiness, and straightforwardness (Leslie & Fleenor, 1998). Regardless of the labels, these competencies are defined in a manner consistent with the dictionary definition of integrity. Integrity-related competencies can be assessed in many ways—structured interviews, background checks, assessment centers, and so forth. Most common may be the 360-degree method of gathering performance ratings from superiors, peers, and subordinates as well as managers themselves. Thus, we focus on ratings as a method for assessing an integrity competency because it is so common in organizational life.[1]

We content-analyzed the items used to rate integrity contained in the Center for Creative Leadership's review of 24 popular assessment instruments (Leslie & Fleenor, 1998). Sample items include "Worthy of trust, believable" (p. 68), "Maintains high standards of personal integrity" (p. 91), "Shows consistency between words and actions" (p. 248), "Tells the truth" (p. 261), "Can be trusted to do what he or she says will be done" (p. 273), and "Is trusted by people in the work group" (p. 285). These items resemble those found in other competency rating instruments commonly used in organizations.

Three points about ratings of an integrity competency should be noted. First, just as leadership research focuses on the positives, so too do instruments for rating the integrity of managers—the items only reflect the desirable end of the integrity construct. None of the items listed above concern *lying, cheating, stealing, deceiving,* or *manipulating*. In fact, we found no items focused on a lack of integrity or unethical behavior in the various instruments we reviewed. The authors of these competency-based rating scales seem to assume that low integrity is defined by the absence of high integrity rather than by the presence of devious behavior. It is not clear how well such items represent the integrity domain.

Second, competency ratings ask observers to rate observed behavior. However, common but minor breaches of integrity (e.g., ignoring commitments, withholding information) can appear to be honest mistakes, not moral violations. Further, serious violations of integrity are usually covert. As the Association of Certified Fraud Examiners (2010, p. 8) observed, "one of the primary characteristics of fraud is that it is clandestine, or hidden; almost all fraud involves the attempted concealment of the crime." Further, the same article reported a study of 1,843 cases where white collar criminals were found guilty; in 86% of the cases, the perpetrator had no prior convictions. Overt and observable managerial misbehavior is a low base-rate phenomenon and most ethical

[1] We thank one of the special issue editors for pointing out that there may be other methods that are more valid and reliable than coworker ratings for measuring integrity. For instance, high-fidelity simulations and strategically designed assessment center exercises may elicit strong indicators of a manager's level of integrity that trained observers can evaluate. However, such methods are expensive and time consuming and are rarely used in practice relative to the competency rating method.

violations are discovered only after the fact. Therefore, ratings of an integrity competency are unlikely to identify managers who have yet to be caught in the act.

Our last point is that integrity ratings are heavily skewed; nearly all managers receive high ratings for integrity—and this does not seem plausible. For example, the CEO of a major corporation with whom we worked believed that integrity was a simple matter: you either have it or you don't. He instituted a performance review that ended with a single dichotomous rating, "This manager conducts business with ethics and integrity." One year every one of his top 400-plus managers received a "Yes" rating. Nonetheless, several of these managers were subsequently indicted for an accounting scandal, bribing a government official, and conspiracy.

Method

We conducted an empirical study of ratings on a competency-based integrity scale that is representative of those commonly used in practice, but with psychometric properties that are perhaps better than is typical. The study had two goals. First, we tested our expectation that few managers would be rated as lacking integrity. Second, to the extent that ratings of integrity failed to identify individuals at the low end of the continuum, we expected that they would not predict managerial performance.

Participants. The data included subordinates' ratings of 672 directors and vice presidents (VPs) employed by a Fortune 500 technology firm. We only included managers and raters based in the United States to control for cross-cultural effects. On average, each manager was rated by five subordinates. For security reasons, precise demographic data were not made available with the research database; however, we are very familiar with the organization and estimate that approximately 75% of managers and 65% of raters were middle-aged white males with a college education. The company has a long tradition of promoting from within and we estimate that most of this sample had worked in the company for five years or more.

Competency measures. Subordinate ratings were collected using a proprietary competency survey developed for the host company. The ratings were gathered for the sole and stated purpose of developmental feedback and all raters were assured that their ratings would be anonymous. Further, they were told that their ratings would be averaged together with other subordinates' ratings, and that those averages would only be reported if three or more subordinate surveys were available. The organization had used this confidentiality policy with similar feedback surveys for several years, and employees were generally familiar with the rules.

The survey included 23 items intended to measure five competencies. Those competencies (and sample items) included *Vision* ("Comes up with a vision for the future"), *Execution* ("Keeps the organization focused on executing the plan"), *Managerial Courage* ("Raises tough issues, even if it makes people uncomfortable"), *Building Talent* ("Develops talent by coaching people"), and *Integrity* (the three items were "Conducts business with integrity," "Tells the truth," and "Puts the organization's interests ahead of his or her personal ambitions"). Raters used a five-point response scale, ranging from "0 = ineffective" to "4 = extremely effective."

We evaluated the adequacy of the measurement model implied by the competency survey using confirmatory factor analysis (CFA). First, we tested how well the rating data fit a five-correlated factors model with five items each for Vision, Execution, Managerial Courage, and Building Talent and the three items for Integrity. The results indicated that the data fit this model well according to Hu and Bentler's (1999) multiple criteria for judging model fit, $\chi^2(220) = 1786.42$, $p < .001$; comparative fit index = .94, root mean square residual = .03, and root mean square error of approximation = .06. In fact, the data fit the five correlated-factors model much better than a one-factor model representing a single, overall "good–bad leader" or "halo" dimension, providing support for the validity of the measurement model representing five distinct, but empirically related, factors.[2]

[2] Full results from the CFAs for the five- and one-factor models are available from the first author upon request.

Finally, we evaluated the degree of interrater reliability and agreement to justify aggregating ratings across subordinates for each focal manager. We used the r_{wg} statistic to determine the degree of interrater agreement (do raters of the same manager provide the same, or very similar, level of ratings?; James, Demaree, & Wolf, 1993) and intraclass correlations (ICC) to determine the degree of interrater reliability (do raters provide ratings that produce the same, or very similar, rank orderings of target managers?). We used *ICC* (1,1) to estimate the interrater reliability of the results from a single rater, and *ICC* (1,5) to estimate the stability of the mean of five raters (Shrout & Fleiss, 1979), since five was the median number of raters per target in our sample. The results are presented in Table 1, and indicate a relatively high degree of interrater agreement (average mean $r_{wg} = .81$) and reliability [mean *ICC* (1,5) = .53] compared to published samples of subordinate ratings (cf. LeBreton, Burgess, Kaiser, Atchley, & James, 2003). Therefore, we computed the average rating across all subordinates for each item and then computed scale scores for each of the five competencies. The internal consistency reliabilities for these aggregate scale scores were high; Cronbach's Coefficient alpha values were .87 for Vision, .88 for Execution, .84 for Managerial Courage, .88 for Building Talent, and .85 for Integrity.

In sum, the foregoing analyses revealed very good psychometric properties for the competency measures and rule out the possibility that poor measurement could complicate the interpretation of our findings.

Criterion measure. We also had an independent measure of overall performance for the directors and VPs. Several months after the competency ratings were made, each director and VP was placed in one of three categories by his or her boss—"highest performing," "solidly performing," and "lowest performing." The organization required a forced distribution so that each of the bosses assigned 30% of their direct reports to the "highest performing" category, 65% to the "solidly performing" category, and 5% to the "lowest performing category." Most of the bosses categorized between eight and 12 direct reports. The bosses then met with their peers to make a case for their classifications. The peers were allowed to challenge the bosses' classifications using facts and specific examples. After a discussion, the peer groups voted on whether to reclassify the subordinate in question. As a result of this challenge process, approximately 10% of managers were moved into different categories; the majority of these changes moved managers into the "solidly performing" category.

The final data set included 28.3% in the highest performing group, 67.3% in the solidly performing group, and 4.3% in the lowest performing group. This forced distribution is arbitrary; it likely produced a restricted range compared to the true underlying performance continuum. Undoubtedly distinctions are lost among the two thirds of the sample in the "solidly performing group." However, the distribution probably accentuates the difference between the highest and the lowest performing groups.

Table 1
Interrater Agreement and Reliability for Subordinate Ratings of Competency Scales

	Interrater Agreement r_{wg}			Interrater Reliability *ICC*	
Competency scale	Median	*M*	*SD*	(1, 1)	(1, 5)
Vision	.86	.83	.15	.20	.55
Execution	.88	.83	.16	.20	.55
Managerial Courage	.86	.82	.17	.17	.50
Building Talent	.82	.77	.21	.15	.47
Integrity	.84	.79	.19	.22	.58

Note. r_{wg} values were computed for each focal manager first, and then descriptive statistics were computed across the sample. *ICC* (1,1) represents the reliability of ratings from a single rater; *ICC* (1,5) represents the stability of the mean of five raters, which is the unit of analysis in subsequent statistical tests.

Results

Table 2 contains the means, standard deviations, and distributions for the five competency scales. As expected, *t* tests indicated that the mean rating for Integrity was significantly higher than the mean rating for every other dimension. The smallest difference occurred between Integrity ($M = 2.68$, $SD = .40$) and Execution ($M = 2.31$, $SD = .42$) where the Integrity score was .90 standard deviations higher, paired $t\ (671) = 22.88$, $p < .001$.

Table 2 shows a second way to consider the elevation of the Integrity ratings. The percentages in the right side of the table indicate the proportion of the total sample scoring between each value on the five-point response scale. For instance, the proportion of the sample rated below the middle anchor, “2 = effective,” was 21.1% for Vision, 19.3% for Execution, 25.3% for Managerial Courage, 35.1% for Building Talent, but only 4.0% for Integrity. Not a single one of the 672 managers in the sample received an Integrity score below “1 = adequate,” presumably the minimally acceptable level. These results are consistent with our claim that competency ratings do not identify managers with integrity issues.

Table 3 presents the correlations among the five competencies. As is typically the case with competency ratings, and as was specified in the measurement model tested with CFA, the scales are correlated with each other. However, Integrity was most highly correlated with Building Talent, which reflects a manager’s concern for subordinates. This is consistent with the definitional position that integrity involves sensitivity to the needs and rights of other people.

Finally, we examined the relationship between subordinates’ competency ratings and bosses’ overall performance rankings. First, we compared the average rating on each competency scale for managers classified as “highest performing,” “solidly performing,” and “lowest performing.” The results are presented in Table 4. Analysis of variance with follow-up Tukey tests for honestly significant differences indicated that the means for Vision, Execution, and Managerial Courage were significantly higher for the highest performing managers compared to the solidly- and lowest performing managers ($p < .001$; partial $\text{eta}^2 = .04$, .04, and .02, respectively). The means for Building Talent and Integrity were not significantly different across the three levels of overall performance. In other words, competency ratings of Integrity did not distinguish the high- from low-performing managers. In fact, there was only a .11 scale-point difference between the average Integrity rating for managers in the “highest performing” and “lowest performing” categories. Further, the mean Integrity rating for “lowest performing” managers was 2.64, which is significantly higher than the mean rating for each of the other four competencies for the “highest performing” managers.

We also used the competency scores as a set of five covariates in a logit regression analysis to predict overall performance. Logit regression is analogous to the familiar linear regression model, except logit regression is more appropriate for criterion variables that are ordinal and categorical like

Table 2
Descriptive Statistics and Distributions of Competency Scale Scores

					Percentage of Sample Scoring between Scale Points			
Competency scale	*M*	*SD*	Min.	Max.	Ineffective–Adequate 0–1	Adequate–Effective 1–2	Effective–Very effective 2–3	Very effective–Extremely effective 3–4
Vision	2.29	(.43)	.70	3.60	0.4%	20.7%	75.7%	3.2%
Execution	2.31	(.42)	.50	3.37	1.0%	18.3%	76.9%	3.8%
Managerial Courage	2.22	(.42)	.70	3.60	0.4%	24.9%	72.3%	2.4%
Building Talent	2.17	(.51)	.40	3.60	2.1%	33.0%	63.0%	1.9%
Integrity	2.68	(.40)	1.29	3.89	0.0%	4.0%	70.9%	25.1%

Note. Min. = minimum; Max. = maximum. Scale points: 0 = Ineffective; 1 = Adequate; 2 = Effective; 3 = Very effective; 4 = Extremely effective.

Table 3
Correlations Among Competency Scales

Competency scale	Vision	Execution	Managerial Courage	Building Talent	Integrity
Vision	**(.87)**				
Execution	.55	**(.86)**			
Managerial Courage	.63	.66	**(.84)**		
Building Talent	.59	.57	.52	**(.88)**	
Integrity	.44	.48	.38	.60	**(.85)**

Note. $N = 672$. Coefficients along the diagonal are coefficient alpha estimates of internal consistency reliability. All correlations significant, $p < .001$.

the overall performance variable in our study (Winship & Mare, 1984). These results are presented in Table 5. The set of five competencies significantly predicted overall performance, $\chi^2(5) = 39.83$, $p < .001$, Nagelkerke Pseudo $R = .27$. However, only Vision and Execution made a significant contribution; once again, the Integrity competency did not predict performance.

Conclusion

The foregoing analysis of subordinate ratings of an integrity competency is consistent with our experience with performance ratings in organizations. Specifically, relative to other competency dimensions, ratings of integrity are always more favorable and they rarely identify managers who may lack integrity. Moreover, managers who got lower scores on integrity were no more likely than those with higher integrity scores to be ranked as low performers overall, and this is at odds with the research showing that integrity is prerequisite for effective leadership. Taken together, the preceding analyses provide empirical support for our contention that competency ratings are unlikely to identify managers with integrity issues.

Study 2: The Dubious Reputation Approach to Integrity

There is an alternative method that may identify managers with potential integrity problems. This method differs from competency ratings in two ways. First, it focuses on the low end of the integrity continuum. Second, it does not rely on ratings of observed behavior; rather, it asks subordinates to estimate the likelihood that managers will engage in unethical behaviors. We refer to this as the *dubious reputation* approach because it involves observer evaluations of the dark side of managers' personalities.

Table 4
Competency Scale Scores for Each Level of Overall Performance Ranking

	Overall Performance Rank					
	Lowest Performing, $n = 29$		Solidly Performing, $n = 452$		Highest Performing, $n = 191$	
Competency scale	*M*	*SD*	*M*	*SD*	*M*	*SD*
Vision	2.15^a	(.44)	2.24^a	(.44)	2.42^b	(.38)
Execution	2.14^a	(.49)	2.26^a	(.43)	2.44^b	(.36)
Managerial Courage	2.10^a	(.51)	2.19^a	(.43)	2.30^b	(.35)
Building Talent	2.03	(.57)	2.14	(.52)	2.26	(.47)
Integrity	2.64	(.49)	2.66	(.41)	2.75	(.37)

Note. Means within a row with a different superscript are significantly different ($p < .01$).

Table 5
Summary of Logit Regression Analysis Using Competency Scales to Predict Overall Performance Ranking

Competency Scale	Wald Parameter
Vision	10.37***
Execution	11.95***
Managerial Courage	1.75
Building Talent	.67
Integrity	.04
Nagelkerke Pseudo *R*	.27***

Note. Wald parameters are analogous to beta-weights in linear regression.
*** $p < .001$.

The previous study indicated how ratings focused on positive behaviors may not identify managers with integrity problems, but ratings focused on the undesirable portion of the domain may. As Craig and Gustafson's (1998) research led them to conclude, "global perceptions of supervisor integrity . . . [are] . . . a function of discrete, and primarily *destructive,* supervisor behaviors" (p. 134, italics added).

Competency measures do not include items from the low end of the integrity continuum for at least three reasons. First, leadership research traditionally emphasizes the positive qualities of leaders (Kellerman, 2004; Padilla et al., 2007). Second, competencies are positive attributes that characterize high performers (Boyatzis, 1982). This is consistent with the tendency of students of leadership to seek "the right stuff" while ignoring "the wrong stuff." A notable exception is research studying managerial derailment (cf. McCall & Lombardo, 1983). A key lesson is that derailment is less about lacking "the right stuff" and more about having "the wrong stuff" (e.g., arrogance and abrasiveness; J. Hogan et al., 2010). Finally, it is more politically correct to ask if managers tell the truth than to ask if they lie, cheat, and steal. But refusing to ask such pointed questions makes it hard to identify managers who lack integrity.

It is difficult to rate low-integrity behaviors because they are hard to observe. For example, employees are rarely caught stealing; however, the base rate of employee theft is about 50% (Wimbush & Dalton, 1997). Although subordinates may not directly observe their manager's misconduct, they can estimate the likelihood that their manager would misbehave when opportunities arise (Craig & Gustafson, 1998).

Moses, Hollenbeck, and Sorcher (1993) argued that data regarding a manager's observed behavior is less useful than data based on how others expect a manager to behave in the future. This is because ratings reflect perceptions of behavior, not actual behavior. As Moses et al., (p. 286) note: "In many cases, the behaviors behind others' perceptions are never clearly delineated. The perceptions are formed as a result of cumulative observations and inferences about an individual. Sometimes these inferences are based on concrete behavioral data. More often they are the collective hypothesis which forms the aura behind a person's reputation." Moreover, a manager's reputation—that is, shared expectations about his or her future behavior—is the primary determinant of how other people respond to that manager.

Asking subordinates to estimate the likelihood that a manager will behave unethically is consistent with this view. People intuitively estimate the likelihood that others are cooperative or selfish (Macy & Skvoretz, 1998). People search for signs of character in others—eye contact in conversations, the company they keep—when deciding whether to trust them (Frank, 1988). Given the vulnerability of their position, subordinates should be acutely alert to cues about a manager's trustworthiness (Van Vugt et al., 2008).

Craig and Gustafson (1998) developed and validated the Perceived Leader Integrity Scale (PLIS), which is the epitome of the dubious reputation method. The scale contains 31 items focused on unethical behaviors. The items reflect both character (e.g., "is vindictive") and bad conduct (e.g., "always gets even"). Some items concern breaking rules (e.g., "would steal from the organization"), and some concern expectations about how a manager would treat the person providing the ratings (e.g., "would allow me to be blamed for his or her mistakes," "would lie to me"). Most of the items ask the rater to estimate the likelihood that a leader would engage in an unethical behavior. Confirmatory factor analysis of data from multiple independent samples indicates that the 31 items cohere in one general factor (Craig & Gustafson, 1998). Furthermore, ratings on the PLIS do not appear to be distorted by impression management; in both a student and a field sample, PLIS ratings were unrelated to responding in a socially desirable manner (Craig & Gustafson, 1998). Also, subordinate ratings on the PLIS are highly correlated with their job satisfaction and desire to quit (Craig & Gustafson, 1998). However, there is no research regarding the extent to which the PLIS identifies managers with low integrity or how it compares with other measures that predict leadership effectiveness.

Method

We conducted a study with a short form of the PLIS to determine whether it would identify managers with integrity issues. We asked two questions. First, will this method place managers at the low end of the integrity continuum? Second, will it predict leadership effectiveness?

Participants. We gathered data from 80 employed MBA students at a university in the southeastern United States. They received extra credit for participating in the study. The participants were 60% male, ranging in age from 24 to 44 ($M = 31.5$ years), with from 2 to 23 years of job experience ($M = 9.2$ years). They reported currently working in a variety of industries, and most reported being in "middle management." We asked participants to rate their job satisfaction and then to rate their supervisor on a series of leadership scales. They reported working for their supervisors between one and 12 years ($M = 3.63$ years); 70% of the supervisors were male.

Measures. Participants rated their supervisors using the Initiating Structure and Consideration scales of the Leader Behavior Description Questionnaire—form XII (Stogdill, 1963). These scales are the most valid measures of the constructs from the two-factor leader behavior paradigm (Judge, Piccolo, & Ilies, 2004). Each scale includes 10 items (sample Initiating Structure item: "lets employees know what is expected of them"; Consideration item: "looks out for the personal welfare of employees"). Cronbach's coefficient alpha for each scale was .91.

We used a short version of the PLIS to represent the dubious reputation approach to Perceived Integrity. A confirmatory factor analysis indicated that this eight-item version of the scale measured a single general factor, as does the full 31-item scale (Craig, n.d.). Further, the short version correlates .95 with the full version (Craig, n.d.). In our sample of ratings on the short version, there was also strong evidence for a single dominant factor (the first eigenvalue was eight times larger than the second, which was only 1.6 times larger than the third) and Cronbach's coefficient alpha was .93.

Participants rated their managers for Initiating Structure, Consideration, and Perceived Integrity on a five-point response scale representing how well the items described their manager, ranging from "0 = not all" to "4 = perfectly."

We also collected two scores that represent effective leadership. We used Brayfield and Rothe's (1951) five-item scale to measure subordinate job satisfaction (sample item, "I find real enjoyment in my work"). We used Tsui's (1984) three-item reputational effectiveness scale to measure the perceived effectiveness of the supervisors (sample item: "This manager absolutely meets my expectations in his or her managerial roles and responsibilities"). Participants rated Job Satisfaction and Perceived Effectiveness using a seven-point response scale ranging from "1 = completely disagree" to "7 = completely agree." Both scales were internally consistent (Cronbach's alpha = .87 for Job Satisfaction, .90 for Perceived Effectiveness). Descriptive statistics were $M = 4.07$, $SD = 1.37$ for Job Satisfaction and $M = 3.96$, $SD = 1.69$ for Perceived Effectiveness.

Results

Higher scores are more desirable on Initiating Structure and Consideration; because the PLIS items describe unethical behavior, higher scores are less desirable. To make the scales comparable, we reverse scored the PLIS scale so that higher scores were more desirable. The means, standard deviations, and distributions for the three leadership scales are presented in Table 6. Similar to our earlier findings, the mean rating for Perceived Integrity was higher than the mean rating for the other two scales. The smallest difference was between Initiating Structure ($M = 2.37$, $SD = .83$) and Perceived Integrity ($M = 2.66$, $SD = .97$); the Perceived Integrity score was .32 standard deviations higher, paired $t(79) = 3.48$, $p < .001$. However, this measure was approximately three times less elevated compared to the competency ratings reported above (.32 *SD*s vs. .90 *SD*s, respectively).

Table 6 also shows the proportion of the total sample scoring between each value on the five-point response scale. The proportion of the sample rated lower than the middle point on the scale, "2," was 30.0% for Initiating Structure, 16.3% for Consideration, and 28.8% for Perceived Integrity. This indicates that a significant number of managers were seen as having potential integrity issues. Thus, the PLIS yielded more variability and a much higher incidence of low scores than the integrity competency scale in the first study.

Table 7 presents the correlations among the three leadership scales and the two indicators of leadership effectiveness. Perceived Integrity was more highly correlated with Consideration than Initiating Structure, which again is consistent with the notion that integrity is primarily concerned with attending to the needs and rights of other people. Next, we examined the relationship between the leadership scales and the effectiveness variables. The results for Initiating Structure and Consideration were consistent with prior research: Job Satisfaction was more correlated with Consideration while Perceived Effectiveness was more correlated with Initiating Structure (Judge et al., 2004). However, both Job Satisfaction and Perceived Effectiveness were more highly correlated with Perceived Integrity than with either Initiating Structure or Consideration.

Finally, we entered the three leadership scales as independent variables in separate regression analyses to predict Job Satisfaction and Perceived Effectiveness. These results are presented in Table 8. The full regression model significantly predicted Job Satisfaction, $F(3, 76) = 16.64$, $p < .001$, $R = .63$. However, only Perceived Integrity made a significant contribution. The full regression model also significantly predicted Perceived Effectiveness, $F(3, 76) = 42.03$, $p < .001$, $R = .79$. Both Perceived Integrity and Initiating Structure made significant contributions to the prediction of Perceived Effectiveness. Consideration added nothing beyond Perceived Integrity to the prediction of either outcome. And in both analyses, Perceived Integrity was the strongest of the three predictors.

Overall Discussion

We have argued that competency ratings of observed integrity behaviors are not a good way to evaluate the integrity of managers because this method does a poor job of identifying managers with potential integrity issues. Instead, we have proposed the "dubious reputation" method based on subordinates' expectations that their manager would behave unethically. The results of two studies

Table 6
Descriptive Statistics and Distributions of Leadership Scale Scores

					Proportion of Sample Scoring Between Scale Points			
Scale	*M*	*SD*	Min.	Max.	Not at all–Barely 0–1	Barely–Somewhat 1–2	Somewhat–Well 2–3	Well–Perfectly 3–4
Initiating Structure	2.37	(.83)	.30	3.90	2.5%	33.8%	41.2%	22.5%
Consideration	2.34	(.58)	.30	3.60	1.3%	15.0%	68.7%	15.0%
Perceived Integrity	2.66	(.97)	.25	4.00	7.5%	21.3%	31.2%	40.0%

Note. Min. = minimum; Max. = maximum. Scale points: 0 = Not at all; 1 = Barely; 2 = Somewhat; 3 = Well; 4 = Perfectly.

Table 7
Correlations Among Leadership Scales and Effectiveness Indicators

	Initiating Structure	Consideration	Perceived Integrity	Job Satisfaction	Perceived Effectiveness
Initiating Structure	**(.91)**				
Consideration	.43	**(.91)**			
Perceived Integrity	.36	.54	**(.93)**		
Job Satisfaction	.42	.47	.59	**(.87)**	
Perceived Effectiveness	.63	.59	.70	.71	**(.90)**

Note. $N = 80$. Coefficients along the diagonal are coefficient alpha estimates of internal consistency reliability. All correlations significant, $p < .001$.

suggest that the dubious reputation approach may indeed be a better way to evaluate the integrity of managers than competency ratings. A measure representative of the dubious reputation method, the PLIS scale, identified a sizable number of managers at the unethical end of the continuum, whereas virtually no manager received a low score on the integrity competency. And the results from the correlation and regression analyses using the PLIS were consistent with prior research suggesting that trust and leader integrity is the primary determinant of employee attitudes and leadership perceptions. The PLIS was by far the strongest of three predictors of subordinate job satisfaction and the perceived effectiveness of their managers. In contrast, ratings on the integrity competency were unrelated to managerial performance, raising serious questions about the validity of this method for measuring integrity.

Limitations

Our studies have limitations that make it difficult to directly compare the dubious reputation method with the competency method of measuring integrity. First, the samples were very different; the competency sample included ratings from over 3,300 subordinates of 672 senior managers from one corporation. The dubious reputation sample included only one subordinate for each of 80 managers from a variety of organizations. It may be that some factor unique to the one corporation may account for the low incidence of low scores on the competency scale (e.g., strong norms against criticizing superiors).

Also, the competency analysis used subordinate ratings of integrity to predict superior evaluations of overall performance, whereas the dubious reputation analysis used subordinate ratings of integrity to predict their own ratings of job satisfaction and perceived effectiveness. Using predictor and outcome measures based on the same rating source likely inflated the relationships in the second study. However, our primary concern was with the predictive validity of the measure of integrity in each study relative to the other measures of leadership in the same study. The dubious reputation

Table 8
Summary of Regression Analyses Using Leadership Scales to Predict Effectiveness Indicators

	Job Satisfaction β	Perceived Effectiveness β
Initiating Structure	.12	.33***
Consideration	.17	.17
Perceived Integrity	.44***	.46***
Model R	.63***	.79***

*** $p < .001$.

approach was the best predictor of leadership in the second study; the competency approach was the poorest predictive measure in the first study. Nonetheless, it is possible that our results would have been different had we used subordinate job attitudes and overall ratings of their bosses in the first study.

Perhaps the biggest difficulty in comparing the two methods concerns how the measures of integrity were computed and intended to be used. In the competency study, we used the average rating across several subordinates. This aggregation reduced the variability and increased the reliability in the integrity scores relative to the dubious reputation study—which used ratings from only one subordinate per manager. This may have accentuated the differences between the two samples in the proportion of managers who scored low on integrity. It is unlikely that this accounts for all of the differences, but it is impossible to quantify the effect with our current dataset. Finally, in the competency study, raters knew their ratings would be provided to their managers as developmental feedback. Although raters were assured of anonymity and the organization had a good track record for maintaining confidentiality, it is possible that they felt less comfortable providing low ratings of integrity. In the dubious reputation study, raters knew that their bosses would not see their ratings. It is well documented that raters give more favorable ratings when they know they will be accountable for them (Curtis, Harvey, & Ravden, 2005), and this may have affected our results to some degree.

Despite these limitations, the results presented here suggest that the dubious reputation approach offers some advantages over the competency approach for identifying managers with potential integrity problems. We regard the present effort as a proof of concept for the dubious reputation approach. Further research is needed in applied settings under typical conditions to confirm the results that we observed in study two. In particular, we need to know if subordinates will provide low ratings of their bosses' integrity when they know the survey results will be presented in some form to those bosses. Further, longitudinal research is needed that compares how well the two methods identify managers who eventually get into ethical trouble.

Implications

Our results have implications for human resource and talent management professionals and consulting psychologists. First, they raise doubts about the ability of typical competency ratings to identify managers who may lack integrity. We suggest that organizations using competency ratings to conduct "integrity audits" consider alternative approaches. We also recommend that coaches and consultants who focus on issues of character and integrity in development rely on more than just competency ratings of integrity.

One possibility is to replace competency ratings with ratings based on the dubious reputation method. For instance, the PLIS scale is in the public domain and available for free at www.sbcraig.com/plis. Another option would be to construct similar scales based on the dubious reputation method by focusing on subordinates' expectations that their boss would behave unethically if given the chance. A third option would be to create hybrid scales that combine the positive items of competency scales with more pointed items from the dubious reputation perspective.[3]

Of course, asking subordinates to speculate on the likelihood that their manager would behave unethically is a process that needs to be implemented with great care. Key issues include anonymity and confidentiality as well as safety mechanisms that encourage candid ratings by protecting subordinates from retribution, a clearly communicated policy on how the data will be used, and visible support from senior leadership and their commitment to a culture of integrity. Finally, collecting such information should be accompanied by a commitment to take action when results suggest that there may be a problem with an unethical manager. And this is where things get complicated. Obviously, it is ill-advised to remove a manager because subordinates believe that the manager might behave unethically. We suggest using this kind of data as a red flag—an indication

[3] We thank an anonymous reviewer for this suggestion.

that there may be a problem that warrants an investigation into the leader, the attitudes of the direct reports, and the performance of the unit.

Finally, if the results are to be fed back to managers, this process would require feedback facilitators to tell managers that their subordinates expect that they will behave unethically, and this is an awkward position in which to be. There is some truth to the view that a good reputation is like virginity: once it is lost, it is hard to get back. Nonetheless, these perceptions exist whether or not they are communicated and it seems important for managers to know if their subordinates mistrust them. The key is to help these managers accept that the perceptions exist and to accept responsibility for addressing them. Developmental work can focus on behaviors that promote trust and communicate that the manager is motivated by factors other than pure self-interest (e.g., showing concern for others and the larger organization, prosocial acts of self-sacrifice).

Closing Thoughts

We would like to emphasize three points in closing. First, managerial incompetence is not the same thing as managerial malfeasance; incompetence comes from a lack of capability, malfeasance comes from a lack of character. Moreover, although the base rate for managerial incompetence has been estimated in at least 12 professional sources to average about 50% (J. Hogan et al., 2010), published estimates of the base rate for flawed managerial integrity are hard to find. It seems that researchers and thoughtful practitioners have treated this topic as taboo. Nonetheless, it is critical to establish guidelines for work in this area to progress. Implicit in our dismal evaluation of competency ratings of integrity is the assumption that a significant proportion of managers are at risk for behaving badly. But opinions seem to vary; for instance, one reviewer of this paper firmly maintained that events like Enron and Tyco are anomalies. On the other hand, ethnographic researchers and corporate ethics experts have suggested that these high-profile cases are just the tip of the iceberg (Jennings, 2006; Sayles & Smith, 2006). Based on our experience with individual assessments of thousands of senior managers, consulting with dozens of major corporations, and on the data reported in Table 6, we estimate the base rate for low integrity managers to be in the 10% to 20% range. This implies that every organization contains a sizable number of managers capable of breeding mistrust and disengagement among employees and harming their organizations. Checks and balances, proper oversight, and periodic assessments are needed to prevent these individuals from acting on their selfish agendas (Kaiser & Hogan, 2007; Padilla et al., 2007).

Second, we believe that the competency rating method seriously underestimates the number of managers with integrity issues. This is because the typical competency survey defines integrity strictly in terms of desirable ethical behaviors that have been observed. The norm is for virtually all managers to receive acceptable ratings on these measures of integrity. Also feeding the problem is that managers are more often evaluated by their bosses than by their subordinates; bosses rarely see the dark side of managers, subordinates routinely see the dark side. Bosses need better methods for learning about how subordinates view their direct supervisor's integrity. Alternative methods for detecting low integrity—such as simulations, assessment centers, enhanced background checks, and perhaps even specially designed interviews—may also be needed, especially for selection purposes.

Our third point is that it is rare for managers to get caught misbehaving. However, managers who are likely to behave unethically let cues slip through their leaky channels, and subordinates use these cues to frame the reputations of their managers. Reputations reflect the likelihood of people behaving in certain ways, and subordinates' estimates of the likelihood that their manager will behave unethically probably come closer to predicting managerial misconduct than any other source of data. This is ultimately an empirical matter and we encourage research, rather than rhetoric and opinion, to evaluate our claim.

References

Allport, G. W. (1937). *Personality: A psychological interpretation.* New York: Henry Holt & Company.

Association of Certified Fraud Examiners. (2010). *2010 Global fraud study: Report to the nations on occupational fraud and abuse.* Austin, TX: Association of Certified Fraud Examiners.

Bass, B. M., & Steidlmeier, P. (1999). Ethics, character, and authentic transformational leadership behavior. *Leadership Quarterly, 10,* 181–217.

Beehr, T. A., Ivanitskaya, L., Hansen, C. P., Erofeev, D., & Gudanowski, D. M. (2001). Evaluation of 360 degree feedback ratings: Relationships with each other and with performance and selection predictors. *Journal of Organizational Behavior, 22,* 775–788.

Blau, P. (1964). *Exchange and power in social life.* New York: Wiley.

Boehm, C. (1999). *Hierarchy in the forest.* London: Harvard University Press.

Boyatzis, R. E. (1982). *The competent manager: A model for effective performance.* New York: Wiley.

Brayfield, A. H., & Rothe, H. F. (1951). An index of job satisfaction. *Journal of Applied Psychology, 35,* 307–311.

Brown, M. E., & Trevino, L. K. (2006). Ethical leadership: A review and future directions. *Leadership Quarterly, 17,* 595–616.

Burns, J. (1978). *Leadership.* New York: Harper & Row.

Chance, M. R. A. (1967). Attention structure as the basis of primate rank order. *Man, 2,* 503–518.

Character. (n.d.). In *Oxford dictionaries*. Retrieved from http://oxforddictionaries.com

Ciulla, J. (2004). Ethics and leadership effectiveness. In J. Antonakis, A. T. Cianciolo, & R. J. Sternberg (Eds.), *The Nature of leadership* (pp. 302–327). Thousand Oaks: Sage.

Cohen, W. A. (2009). *Drucker on leadership.* San Francisco: Jossey-Bass.

Cosmides, L., & Tooby, J. (1989). Evolutionary psychology and the generation of culture, Part II. Case study: A computational theory of social exchange. *Ethology and Sociobiology, 10,* 51–97.

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. Barkow, L. Cosmides, & J. Tooby (Eds.) *The adapted mind* (pp. 163–28). New York: Oxford University Press.

Craig, S. B. (n.d.). A short form of the Perceived Leader Integrity Scale (PLIS). Retrieved from http://www.sbcraig.com/plis/short.htm

Craig, S. B., & Gustafson, S. B. (1998). Perceived leader integrity scale. *Leadership Quarterly, 9,* 127–145.

Curtis, A. B., Harvey, R. D., & Ravden, D. (2005). Sources of political distortions in performance appraisal. *Group & Organization Management, 30,* 42–60.

Dansereau, F., Graen, G., & Hagan, W. (1975). A vertical dyadic linkage approach to leadership within formal organizations. *Organizational Behavior and Human Performance, 13,* 46–70.

Den Hartog, D. N., House, R. J., Hanges, P. J., Ruiz-Quintanilla, S. A., & Dorfman, P. W. (1999). Culture-specific and cross-culturally generalizable implicit leadership theories: A longitudinal investigation. *Leadership Quarterly, 10,* 219–256.

Dirks, K. T., & Ferrin, D. L. (2002). Trust in leadership: Meta-analytic findings and implications for research and practice. *Journal of Applied Psychology, 87,* 611–628.

Ethics. (n.d.). In *Oxford dictionaries*. Retrieved from http://oxforddictionaries.com

Frank, R. (1988). *Passions within reason: The strategic role of the emotions.* New York: Norton.

Freud, S. (1921). Group psychology and the analysis of the ego. In J. Strachey (Ed. and Trans.), *The standard ed. of the complete works of Sigmund Freud, Vol 19.* London: Hogarth Press.

Gerstner, C. R., & Day, D. V. (1997). Meta-analytic review of leader-member exchange theory: Correlates and construct issues. *Journal of Applied Psychology, 82,* 827–844.

Greenberg, J. (1990). Employee theft as a response to underemployment inequity: The hidden costs of pay cuts. *Journal of Applied Psychology, 75,* 561–568.

Heidemeier, H., & Moser, K. (2009). Self-other agreement in job performance ratings: A meta-analytic test of a process model. *Journal of Applied Psychology, 94,* 353–370.

Hogan, J., Hogan, R., & Kaiser, R. B. (2010). Management derailment: Personality assessment and mitigation. In S. Zedeck (Ed.) *American Psychological Association handbook of industrial and organizational psychology, vol.* 3 (pp. 555–575). Washington, DC: American Psychological Association.

Hogan, J., & Ones, D. S. (1997). Conscientiousness and integrity tests at work. In R. Hogan, J. A. Johnson, and S. R. Briggs (Eds.) *Handbook of personality psychology* (pp. 849–870). San Diego, CA: Academic Press.

Hogan, R. (2007). *Personality and the fate of organizations.* Hillsdale, NJ: Erlbaum.

Hogan, R., Curphy, G. J., & Hogan, J. (1994). What we know about leadership: Personality and effectiveness. *American Psychologist, 49,* 493–504.

Hogan, R., & Hogan, J. (2001). Assessing leadership: A view of the dark side. *International Journal of Selection and Assessment, 9,* 40–51.

Hogan, R., & Kaiser, R. B. (2005). What we know about leadership. *Journal of General Psychology, 9,* 169–180.

Howard, A., & Thomas, J. N. (2010). Executive and managerial assessment. In J. C. Scott & D. H. Reynolds (Eds.), *Handbook of workplace assessment* (pp. 395–436). San Francisco: Jossey-Bass.

Howell, J. M., & Avolio, B. J. (1992). The ethics of charismatic leadership: Submission or liberation? *Academy of Management Executive, 6,* 43–54.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure modeling: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1–55.

Integrity. (n.d.). In *Oxford dictionaries*. Retrieved from http://oxforddictionaries.com

James, L. J., Demaree, R. G., & Wolf, G. (1993). r_{wg}: An assessment of within-group interrater agreement. *Journal of Applied Psychology, 78,* 306–309.

Jennings, M. M. (2006). *The seven signs of ethical collapse.* New York: St. Martin's Press.

Judge, T. A., Piccolo, R. F., & Ilies, R. (2004). The forgotten ones? The validity of consideration and initiating structure in leadership research. *Journal of Applied Psychology, 89,* 36–51.

Kaiser, R. B., & Hogan, R. (2007). The dark side of discretion. In R. Hooijberg, J. Hunt, J. Antonakis, & K. Boal (Eds.), *Being there even when you are not: Leading through strategy, systems and structures* (pp. 177–197). London: Elsevier Science.

Kellerman, B. (2004). *Bad leadership.* Boston: Harvard Business School Press.

LeBreton, J. M., Burgess, J. R. D., Kaiser, R. B., Atchley, E. K., & James, L. R. (2003). The restriction of variance hypothesis and interrater reliability and agreement: Are ratings from multiple sources really dissimilar? *Organizational Research Methods, 6,* 78–126.

Leslie, J. B., & Fleenor, J. W. (1998). *Feedback to managers: A review and comparison of multi-rater instruments for management development.* Greensboro, NC: Center for Creative Leadership.

Lord, R., Foti, R. J., & De Vader, C. L. (1984). A test of leadership categorization theory. *Organizational Behavior and Human Performance, 34,* 343–378.

Lord, R. G., & Maher, K. J. (1993). *Leadership and information processing: Linking perceptions and performance.* London: Routledge.

MacKinnon, D. W. (1944). The structure of personality. In J. McVicker Hunt (Ed.), *Personality and the behavior disorders* (Vol. 1, pp. 3–48). New York: Ronald Press.

Macy, M. W., & Skvoretz, J. (1998). The evolution of trust and cooperation between strangers: A computational model. *American Sociological Review, 63,* 638–660.

Maner, J. K., & Mead, N. L. (2010). The essential tension between leadership and power: When leaders sacrifice group goals for the sake of self-interest. *Journal of Personality and Social Psychology, 99,* 482–497.

McCall, M. W., Jr., & Lombardo, M. M. (1983). *Off the track: Why and how successful executives get derailed.* Tech. Rep. No. No. 21. Greensboro, NC: Center for Creative Leadership.

Mischel, W. (1977). The interaction of person and situation. In D. Magnusson & N. S. Endler (Eds.), *Personality at the crossroads: Current issues in interactional psychology* (pp. 333–352). Hillsdale, NJ: Erlbaum.

Morals. (n.d.). In *Oxford dictionaries*. Retrieved from http://oxforddictionaries.com

Moses, J., Hollenbeck, G. P., & Sorcher, M. (1993). Other people's expectations. *Human Resource Management, 32,* 283–297.

Murphy, K. R., & Cleveland, J. N. (1995). *Understanding performance appraisal: Social, organizational, and goal-based perspectives.* Thousand Oaks, CA: Sage.

Northouse, P. (2006). *Leadership: Theory and practice* (4th ed.). London: Sage.

O'Connor, J., Mumford, M., Clifton, T., Gessner, T., & Connelly, M. (1995). Charismatic leaders and destructiveness: An historiometric study. *Leadership Quarterly, 6,* 529–555.

Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (1993). Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance. *Journal of Applied Psychology, 78,* 679–703.

Padilla, A., Hogan, R., & Kaiser, R. B. (2007). The toxic triangle: Destructive leaders, vulnerable followers, and conducive environments. *Leadership Quarterly, 18,* 176–194.

Parry, K., & Proctor-Thomson, S. B. (2002). Perceived integrity of transformational leaders in organisational settings. *Journal of Business Ethics, 35,* 75–96.

Piaget, J. (1965). *The moral judgment of the child.* New York: The Free Press.

Pojman, L. P. (1995). *Ethical theory: Classical and contemporary readings* (2nd ed.). Belmont, CA: Wadsworth.

Sayles, L. R., & Smith, C. J. (2006). *The rise of the rogue executive: How good companies go bad and how to stop the destruction.* Upper Saddle River, NJ: Prentice Hall.

Shockley-Zalabak, P., Ellis, K., & Winograd, G. (2000). Organizational trust: What it means, why it matters. *Organization Development Journal, 18,* 35–48.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86,* 420–428.

Stamoulis, D. (2009). *Senior executive assessment: A Key to responsible corporate governance.* Malden, MA: Wiley-Blackwell.

Stogdill, R. M. (1963). *Manual for the Leader Behavior Description Questionnaire, Form XII.* Columbus, OH: Bureau of Business Research, Ohio State University.

Thompson, A. D., Grahek, M., Phillips, R. E., & Fay, C. L. (2008). The search for worthy leadership. *Consulting Psychology Journal: Practice and Research, 60,* 366–382.

Tsui, A. S. (1984). A role-set analysis of managerial reputation. *Organizational Behavior and Human Performance, 34,* 64–96.

Van Vugt, M., Hogan, R., & Kaiser, R. B. (2008). Leadership, followership, and evolution: Some lessons from the past. *American Psychologist, 63,* 182–196.

Wimbush, J. C., & Dalton, D. R. (1997). Base rate for employee theft: Convergence of multiple methods. *Journal of Applied Psychology, 82,* 756–763.

Winship, C., & Mare, R. D. (1984). Regression models with ordinal variables. *American Sociological Review, 49,* 512–525.

Received June 6, 2010
Revision received November 5, 2010
Accepted November 10, 2010 ■